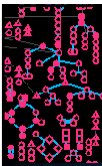




## A Pathway/Genome Database for *Caulobacter crescentus*

**Peter D. Karp, Ph.D.**  
Bioinformatics Research Group  
SRI International  
pkarp@ai.sri.com

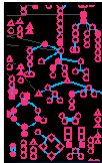
<http://www.ai.sri.com/pkarp/>  
<http://ecocyc.org/>



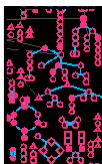
## *Microbial Cell Project*

SRI International  
Bioinformatics

- **SRI International**
  - Peter Karp
  - Pedro Romero
- **Stanford University**
  - Harley McAdams
  - Lucy Shapiro
  - William Lee
  - Allison Hottes
  - Amit Puyani

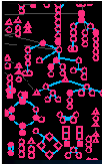


## *C. crescentus Cell Cycle Gene Expression*



## *Software for Network Analysis of Genomes*

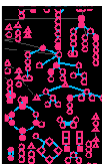
- Pathway Tools software
- Pathway/Genome Databases (PGDBs)
- PGDB for *Caulobacter Crescentus*



## *Pathway Tools Software*

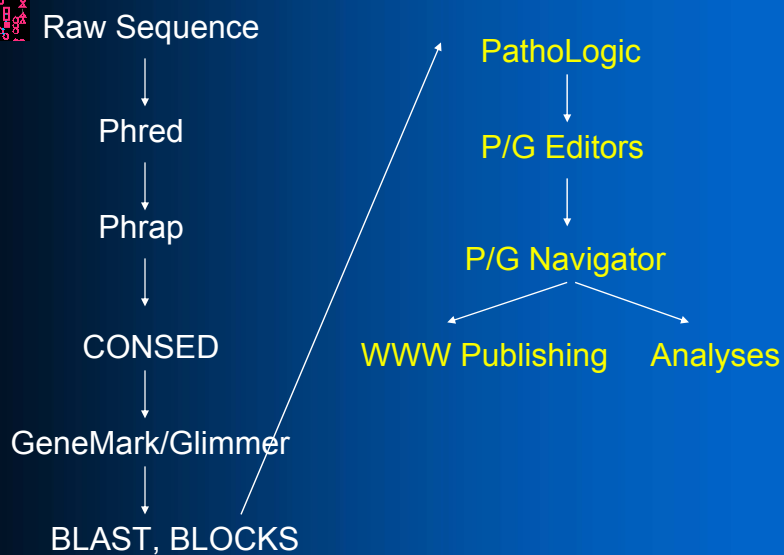
SRI International  
Bioinformatics

- **PathoLogic**
  - Prediction of metabolic network from genome
  - Computational creation of new Pathway/Genome Databases
- **Pathway/Genome Editors**
  - Distributed curation of genome annotations
  - Distributed object database system
  - Interactive editing tools for pathways, enzymes, genes, operons, etc.
- **Pathway/Genome Navigator**
  - WWW publishing of PGDBs
  - Graphic depictions of pathways, chromosomes, operons, regulons
  - Analysis operations
    - ◆ Pathway visualization of gene-expression data
    - ◆ Global comparisons of metabolic networks



## *Sequence Project Workflow*

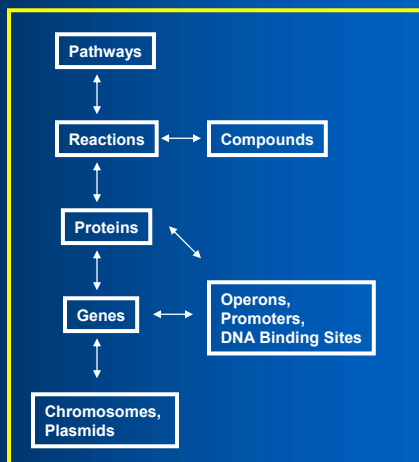
SRI International  
Bioinformatics



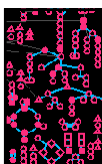


## Pathway/Genome Database

SRI International  
Bioinformatics



CELL



## Pathway/Genome DBs at the SRI Web Site

SRI International  
Bioinformatics

### Literature-based Datasets:

- *MetaCyc*
- *Escherichia coli (EcoCyc)*

### Computationally Derived Datasets:

- *Agrobacterium tumefaciens*
- *Caulobacter crescentus*
- *Chlamydia trachomatis*
- *Bacillus subtilis*
- *Helicobacter pylori*
- *Haemophilus influenzae*
- *Mycobacterium tuberculosis*
- *Mycoplasma pneumonia*
- *Pseudomonas aeruginosa*
- *Saccharomyces cerevisiae*
- *Treponema pallidum*

<http://ecocyc.org/>



## Pathway/Genome DBs Created by External Users

SRI International  
Bioinformatics

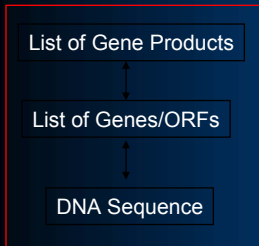
- ***Plasmodium falciparum*, Stanford University**
  - [plasmocyc.stanford.edu](http://plasmocyc.stanford.edu)
- ***Arabidopsis thaliana*, Carnegie Institution of Washington**
- ***Synechocystis*, Carnegie Institution of Washington**
- **Seven other PGDBs in progress by 5 other users**



## Inference of Metabolic Pathways

SRI International  
Bioinformatics

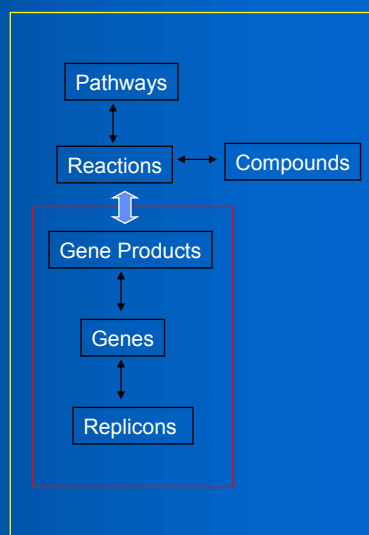
ANNOTATED GENOME  
Structured ASCII Text File

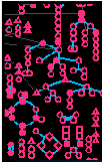


MetaCyc

PathoLogic

Pathway/Genome Database





## MetaCyc Overview

- **Metabolic Encyclopedia**
- **445 pathways, 1115 enzymes, 4218 reactions**
  - 173 *E. coli* pathways; 158 organisms
  - 2381 citations
- **Literature-based DB with extensive references and commentary**
- **Pathways, reactions, enzymes, substrates**

### Genbank Format :

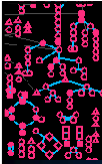
```

gene          422054..423490
               /gene="aroE"
CDS           422054..423490
               /gene="aroE"
               /label="CT370"
               /product="Shikimate 5-Dehydrogenase"
               /db_xref="PID:g3328794"
  
```

### PathoLogic Format :

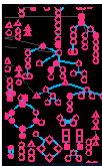
```

ID           CT370
NAME          aroE
STARTBASE    422054
ENDBASE      423490
PRODUCT      Shikimate 5-Dehydrogenase
DBLINK        PID:g3328794
PRODUCT-TYPE P
EC            1.1.1.25
  
```



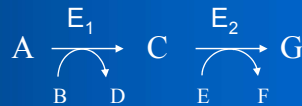
## *PathoLogic Analysis Phases*

- **Trial parsing of input data files**
- **Automated build of initial PGDB**
  - Initialize schema of new PGDB
  - Create DB objects for chromosomes, genes, proteins
  - Assign enzymes to metabolic reactions
  - Import pathways from MetaCyc
- **Define protein complexes**
- **Evaluate pathway evidence**
- **Define metabolic overview diagram**

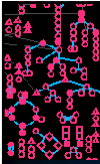


## *Assignment of Enzymes to Pathways*

- **Reactions are the links between enzymes and pathways**
- **Determine biochemical reaction catalyzed by each enzyme**



- **Example enzymes:**
  - ferrochelatase
  - glutamate 1-semialdehyde 2,1-aminomutase
  - porphobilinogen deaminase
- **Match by EC number (142/850 *E. coli*)**
- **Exact enzyme name matching**
- **Inexact enzyme name matching**
- **Literature research**



## Comparison to KEGG Pathway Prediction for *C. crescentus*

SRI International  
Bioinformatics

- **CauloCyc reaction assignments: 527**
- **KEGG Caulobacter reaction assignments: 419**
- **Pathways missed by KEGG analysis:**
  - pyridoxal 5'-phosphate biosynthesis, isopentenyl diphosphate biosynthesis - mevalonate-independent, mannose catabolism, ribosylnicotinamide degradation, aldoxime metabolism, deoxyribonucleotide metabolism
- **KEGG does not make decisions as to presence of pathways**
- **KEGG does not support per-organism editing of pathways**



## Metabolic Comparison of *C. crescentus* to *A. tumefaciens*

SRI International  
Bioinformatics

- **Total pathways:**
  - *C. crescentus*: 130    *A. tumefaciens*: 156
- **Shared pathways: 129**
- **Pathways present in AT but not CC:**
  - arginine glutamate degradation, arginine proline degradation, betaine biosynthesis III, creatinine formate catabolism, D-galactarate catabolism, D-glucarate catabolism, denitrification pathway, enterobacterial common antigen biosynthesis, galactitol catabolism, galactonate catabolism, glucuronate degradation (to xylulose 5-phosphate), glycogen catabolism, heterofermentative lactate fermentation, NAD biosynthesis I, NAD phosphorylation and dephosphorylation, nopaline degradation, octopine degradation, polyisoprenoid biosynthesis, proline utilization 2, pyridine nucleotide cycling, quinate degradation, ribitol degradation, stachydrine catabolism, trypanothione biosynthesis, ureide degradation, xylose catabolism